Spark 3.3.1   # Spark Master at spark://master:7077

URL: spark://master:7077
Alive Workers: 6
Cores in use: 28 Total, 0 Used
Memory in use: 59.3 GiB Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 192 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

▾ Workers (6)

Worker Id	Address	State	Cores	Memory	Resources
worker-20230413105822-147.96.20.35-34415	147.96.20.35:34415	ALIVE	8 (0 Used)	30.1 GiB (0.0 B Used)	
worker-20230413113149-192.168.135.11-39281	192.168.135.11:39281	ALIVE	4 (0 Used)	6.6 GiB (0.0 B Used)	
worker-20230413113149-192.168.135.14-36307	192.168.135.14:36307	ALIVE	4 (0 Used)	6.6 GiB (0.0 B Used)	
worker-20230413113149-192.168.135.15-41663	192.168.135.15:41663	ALIVE	4 (0 Used)	6.6 GiB (0.0 B Used)	
worker-20230413113149-192.168.135.21-34617	192.168.135.21:34617	ALIVE	4 (0 Used)	6.6 GiB (0.0 B Used)	
worker-20230413113150-192.168.135.23-42655	192.168.135.23:42655	ALIVE	4 (0 Used)	2.7 GiB (0.0 B Used)	

▾ Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration

▾ Completed Applications (192)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20230505184356-0271	pyspark-shell	28	1024.0 MiB		2023/05/05 18:43:56	jcaste05	FINISHED	4 s
app-20230505184152-0270	pyspark-shell	28	1024.0 MiB		2023/05/05 18:41:52	jcaste05	FINISHED	4 s
app-20230505183626-0269	pyspark-shell	28	1024.0 MiB		2023/05/05 18:36:26	jcaste05	FINISHED	4 s

---

Archivo Acciones Editar Vista Ayuda

| jcaste05@wild: ~/Ejercicio_Quijote × | | prpa@prpa: ~ × |

```
-rw-r--r--   2 jcaste05 supergroup    2178967 2023-05-05 18:34 /user/jcaste05/quijote.txt
jcaste05@wild:~/Ejercicio_Quijote$ python3 coger_porcentaje_lineas.py /user/jcaste05/quijote.txt quijote_s05_hdfs.txt 0.5
2023-05-05 18:41:51,305 INFO spark.SparkContext: Running Spark version 3.3.1
2023-05-05 18:41:51,376 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2023-05-05 18:41:51,463 INFO resource.ResourceUtils: ==============================================================
2023-05-05 18:41:51,464 INFO resource.ResourceUtils: No custom resources configured for spark.driver.
2023-05-05 18:41:51,464 INFO resource.ResourceUtils: ==============================================================
2023-05-05 18:41:51,464 INFO spark.SparkContext: Submitted application: pyspark-shell
2023-05-05 18:41:51,480 INFO resource.ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> name: memory, amount: 1024, script: , vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
2023-05-05 18:41:51,490 INFO resource.ResourceProfile: Limiting resource is cpu
2023-05-05 18:41:51,490 INFO resource.ResourceProfileManager: Added ResourceProfile id: 0
2023-05-05 18:41:51,532 INFO spark.SecurityManager: Changing view acls to: jcaste05
2023-05-05 18:41:51,532 INFO spark.SecurityManager: Changing modify acls to: jcaste05
2023-05-05 18:41:51,533 INFO spark.SecurityManager: Changing view acls groups to:
2023-05-05 18:41:51,533 INFO spark.SecurityManager: Changing modify acls groups to:
2023-05-05 18:41:51,534 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(jcaste05); groups with view permissions: Se
t(); users  with modify permissions: Set(jcaste05); groups with modify permissions: Set()
2023-05-05 18:41:51,736 INFO util.Utils: Successfully started service 'sparkDriver' on port 38957.
2023-05-05 18:41:51,755 INFO spark.SparkEnv: Registering MapOutputTracker
2023-05-05 18:41:51,779 INFO spark.SparkEnv: Registering BlockManagerMaster
2023-05-05 18:41:51,791 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
2023-05-05 18:41:51,792 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
2023-05-05 18:41:51,794 INFO spark.SparkEnv: Registering BlockManagerMasterHeartbeat
2023-05-05 18:41:51,811 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-66f44944-408e-4e73-96e8-edc786277288
2023-05-05 18:41:51,823 INFO memory.MemoryStore: MemoryStore started with capacity 434.4 MiB
2023-05-05 18:41:51,834 INFO spark.SparkEnv: Registering OutputCommitCoordinator
2023-05-05 18:41:51,869 INFO util.log: Logging initialized @1641ms to org.sparkproject.jetty.util.log.Slf4jLog
2023-05-05 18:41:51,958 INFO server.Server: jetty-9.4.48.v20220622; built: 2022-06-21T20:42:25.880Z; git: 6b67c5719d1f4371b33655ff2d047d24e171e49a; jvm 11.0.18+10-post-Debian-1deb11u1
2023-05-05 18:41:51,971 INFO server.Server: Started @1744ms
2023-05-05 18:41:51,997 INFO server.AbstractConnector: Started ServerConnector@94d803{HTTP/1.1, (http/1.1)}{0.0.0.0:4040}
2023-05-05 18:41:51,997 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
2023-05-05 18:41:52,020 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@72c02b89{/,null,AVAILABLE,@Spark}
2023-05-05 18:41:52,169 INFO client.StandaloneAppClient$ClientEndpoint: Connecting to master spark://192.168.135.1:7077...
2023-05-05 18:41:52,214 INFO client.TransportClientFactory: Successfully created connection to /192.168.135.1:7077 after 21 ms (0 ms spent in bootstraps)
2023-05-05 18:41:52,278 INFO cluster.StandaloneSchedulerBackend: Connected to Spark cluster with app ID app-20230505184152-0270
2023-05-05 18:41:52,281 INFO client.StandaloneAppClient$ClientEndpoint: Executor added: app-20230505184152-0270/0 on worker-20230413105822-147.96.20.35-34415 (147.96.20.35:34415) with 8 cor
e(s)
2023-05-05 18:41:52,282 INFO cluster.StandaloneSchedulerBackend: Granted executor ID app-20230505184152-0270/0 on hostPort 147.96.20.35:34415 with 8 core(s), 1024.0 MiB RAM
2023-05-05 18:41:52,282 INFO client.StandaloneAppClient$ClientEndpoint: Executor added: app-20230505184152-0270/1 on worker-20230413113150-192.168.135.23-42655 (192.168.135.23:42655) with 4
 core(s)
```

Archivo Acciones Editar Vista Ayuda

| jcaste05@wild: ~/Ejercicio_Quijote × | | prpa@prpa: ~ × |

```
2023-05-05 18:41:57,293 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-9791e34f-57e8-4b30-9d3d-89eb1014f55d/pyspark-d619f745-a887-4b03-a0bf-e2a3b579f4ec
2023-05-05 18:41:57,295 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-aefc1b7c-2e79-410f-a953-7cebba6db795
2023-05-05 18:41:57,296 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-9791e34f-57e8-4b30-9d3d-89eb1014f55d
^C
jcaste05@wild:~/Ejercicio_Quijote$ ls
coger_porcentaje_lineas.py  contar_palabras.py  out_hdfs.txt  out_quijote_s05.txt  out_quijote.txt  prueba.txt  quijote_s05_hdfs.txt  quijote_s05.txt  quijote.txt
jcaste05@wild:~/Ejercicio_Quijote$ hdfs dfs -put quijote_s05_hdfs.txt
jcaste05@wild:~/Ejercicio_Quijote$ hdfs dfs -ls /user/jcaste05/
Found 3 items
-rw-r--r--   2 jcaste05 supergroup         44 2023-05-05 18:37 /user/jcaste05/out_hdfs.txt
-rw-r--r--   2 jcaste05 supergroup    2178967 2023-05-05 18:34 /user/jcaste05/quijote.txt
-rw-r--r--   2 jcaste05 supergroup    1055722 2023-05-05 18:42 /user/jcaste05/quijote_s05_hdfs.txt
jcaste05@wild:~/Ejercicio_Quijote$ python3 contar_palabras.py /user/jcaste05/quijote_s05_hdfs.txt out_s05_hdfs.txt
2023-05-05 18:43:55,476 INFO spark.SparkContext: Running Spark version 3.3.1
2023-05-05 18:43:55,545 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2023-05-05 18:43:55,648 INFO resource.ResourceUtils: ==============================================================
2023-05-05 18:43:55,649 INFO resource.ResourceUtils: No custom resources configured for spark.driver.
2023-05-05 18:43:55,649 INFO resource.ResourceUtils: ==============================================================
2023-05-05 18:43:55,649 INFO spark.SparkContext: Submitted application: pyspark-shell
2023-05-05 18:43:55,664 INFO resource.ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> name: memory,
amount: 1024, script: , vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
2023-05-05 18:43:55,674 INFO resource.ResourceProfile: Limiting resource is cpu
2023-05-05 18:43:55,674 INFO resource.ResourceProfileManager: Added ResourceProfile id: 0
2023-05-05 18:43:55,714 INFO spark.SecurityManager: Changing view acls to: jcaste05
2023-05-05 18:43:55,714 INFO spark.SecurityManager: Changing modify acls to: jcaste05
2023-05-05 18:43:55,715 INFO spark.SecurityManager: Changing view acls groups to:
2023-05-05 18:43:55,715 INFO spark.SecurityManager: Changing modify acls groups to:
2023-05-05 18:43:55,715 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(jcaste05); groups with view permissions: Se
t(); users  with modify permissions: Set(jcaste05); groups with modify permissions: Set()
2023-05-05 18:43:55,931 INFO util.Utils: Successfully started service 'sparkDriver' on port 39249.
2023-05-05 18:43:55,951 INFO spark.SparkEnv: Registering MapOutputTracker
2023-05-05 18:43:55,975 INFO spark.SparkEnv: Registering BlockManagerMaster
2023-05-05 18:43:55,987 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
2023-05-05 18:43:55,988 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
2023-05-05 18:43:55,990 INFO spark.SparkEnv: Registering BlockManagerMasterHeartbeat
2023-05-05 18:43:56,005 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-0792f357-9e1f-493b-9a6b-d5caaea814e0
2023-05-05 18:43:56,018 INFO memory.MemoryStore: MemoryStore started with capacity 434.4 MiB
2023-05-05 18:43:56,030 INFO spark.SparkEnv: Registering OutputCommitCoordinator
2023-05-05 18:43:56,062 INFO util.log: Logging initialized @1695ms to org.sparkproject.jetty.util.log.Slf4jLog
2023-05-05 18:43:56,143 INFO server.Server: jetty-9.4.48.v20220622; built: 2022-06-21T20:42:25.880Z; git: 6b67c5719d1f4371b33655ff2d047d24e171e49a; jvm 11.0.18+10-post-Debian-1deb11u1
2023-05-05 18:43:56,156 INFO server.Server: Started @1789ms
```